

ARMY RESEARCH LABORATORY



Analysis of Dual-Band Infrared Imagery from the Multidomain Smart Sensor Field Test

A. Goldberg, T. Fisher, S. Kennerly, S. Der, A. Chan, M. Lander (ARL), C.
Garvin, S. Wang, M. Falco, D. Campagna, and A. Costillo (Sanders)

ARL-TR-996

June 2002

Approved for public release; distribution unlimited.

20021016 197

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Adelphi, MD 20783-1197

ARL-TR-996

June 2002

Analysis of Dual-Band Infrared Imagery from the Multidomain Smart Sensor Field Test

A. Goldberg, T. Fisher, S. Kennerly, S. Der, A. Chan, M. Lander (ARL), C. Garvin, S. Wang, M. Falco, D. Campagna, and A. Costillo (Sanders)
Sensors and Electron Devices Directorate

Abstract

Included is dual-band infrared image data collected as part of the Multi-domain Smart Sensor effort at the U. S. Army Research Laboratory. The ultimate goal of this effort is to produce large format, staring focal plane arrays that are able to see the battlefield in both the 3 to 5 μm (midwave infrared) and 8 to 12 μm (longwave infrared) atmospheric transmission windows. The image data were collected using separate boresighted cameras with equal pixel formats and fields of view during field tests that were conducted during the summer of 1998. This work shows a number of scenarios under which the imagery from one band is superior to that from the other band and various image fusion techniques that help to enhance the visibility of targets. Discussed is a technique for using computer hardware to do the image fusion in real time as well as results of the application of aided target recognition algorithms to the data.

Contents

Introduction	5
Experiment	6
Results and Discussion	10
a. Smoke.....	10
b. Ambient Conditions: Ground Fog and Rain.....	11
c. Image Fusion	12
d. ATR.....	17
1. PCA	18
2. Clutter Rejection	23
3. Experimental Results.....	25
e. Hardware Implementation of Image Fusion.....	30
Current Status and Future Plans	34
Summary and Conclusion	36
Acknowledgments	37
References	39
List of Acronyms	41
Documentation page	43

Figure 1. The Multi-Domain Smart Sensor combines several sensor types with advanced signal processing and aided target recognition for faster and more accurate battlefield threat classification.....	5
Figure 2. Schematic diagram of the camera/image acquisition system for MDSS field test.....	7
Figure 3. Boresighted IR cameras used in the MDSS field tests..	9
Figure 4. Effect of HC smoke on imagery in the visible (a), MWIR (b), and LWIR (c).....	10
Figure 5. Effect of HC smoke on imagery in the MWIR (a) and LWIR (b).....	10
Figure 6. Images of the Ft. A. P. Hill dropzone taken before sunrise under conditions of heavy ground fog.....	11
Figure 7. MWIR (a) and LWIR (b) Images of an M113 (left) and an M2 (right) taken at night under rainy conditions; the range to the targets was 2.1 km	12
Figure 8. MWIR (a) and LWIR (b) Images of an M2 (right) taken at night under rainy conditions	12
Figure 9. IR images of a Bradley Fighting Vehicle (M2) in the MWIR (a) and LWIR (b).....	13
Figure 10. Red-cyan color scheme for fusing MWIR and LWIR images.....	15
Figure 11. Color fusion algorithm applied to the MWIR and LWIR images shown in fig. 6.....	15
Figure 12. Fused IR imagery of an M60 tank through HC smoke (individual MWIR and LWIR images	16
Figure 13. Detection rates on the training and test sets, as a function of false alarms per frame	18
Figure 14. First 50 most dominant PCA eigenvectors for the targets (top 5 rows) and clutter (bottom 5 rows) in the training set	20
Figure 15. First 50 most dominant EST eigenvectors associated with positive (top 5 rows) and negative (bottom 5 rows) eigenvalues for the training set.....	22
Figure 16. Rapid attenuation of eigenvalues in PCA and EST transforms	23
Figure 17. A simple MLP with two layers of weights and shortcut connections	24
Figure 18. Block diagram of the three-processor, CSSP fusion algorithm.....	31
Figure 19. The left image was taken with a cooled MWIR sensor, the	

center image was acquired with an uncooled LWIR sensor and the right is the result of processing with three CSSP to produce a false color enhanced image	31
Figure 20. Image of a man holding a glass filter in front of a lit butane lighter taken with the QWIP dual-band FPA.....	35
Figure 21. Result of image fusion on the images shown in fig. 20.....	35
Table 1. Infrared cameras used in 1998 MDSS field tests	7
Table 2. Target vehicles and their ranges from the camera position	9
Table 3. The number of training/testing image clips used for the clutter-rejection study.	26
Table 4. Hit rates of PCA-d2b_L (LLM chips detected by 2 bands: LLM+LBM) at 10 percent FA rate.....	27
Table 5. Hit rates of EST-d2b_L (LLM chips detected by 2 bands: LLM+LBM) at 10 percent FA rate.....	27
Table 6. Hit rates of PCA-d2b_M (MLM chips detected by 2 bands: MLM+MBM) at 10 percent FA rate	28
Table 7. Hit rates of EST-d2b_M (MLM chips detected by 2 bands: MLM+MBL) at 10 percent FA rate	28
Table 8. Hit rates of PCA-mrg (merged 2 bands: LLMMBL+MLMLBM) at 10 percent FA rate	29
Table 9. Hit rates of EST-mrg (merged 2 bands: LLMMBL+MLMLBM) at 10% false alarm rate	29
Table 10. Hit rates of PCA-d2b_LM (separate eigenvector sets, joint MLP) at 10 percent FA rate	30
Table 11. Hit rates of EST-d2b_LM (separate eigenvector sets, joint MLP) at 10 percent FA rate	30

Introduction

Recently, the U. S. Army Research Laboratory (ARL), in federation with several industry and academic partners, has developed the concept of the Multi-Domain Smart Sensor (MDSS) [i]. This system, shown schematically in figure 1, is envisioned as a single unit combining both passive and active sensor components with advanced signal processing and aided target recognition (ATR) tools. Such a sensor would enhance situational awareness on the battlefield in all ambient conditions by locating and classifying threats with increased effectiveness over existing systems.

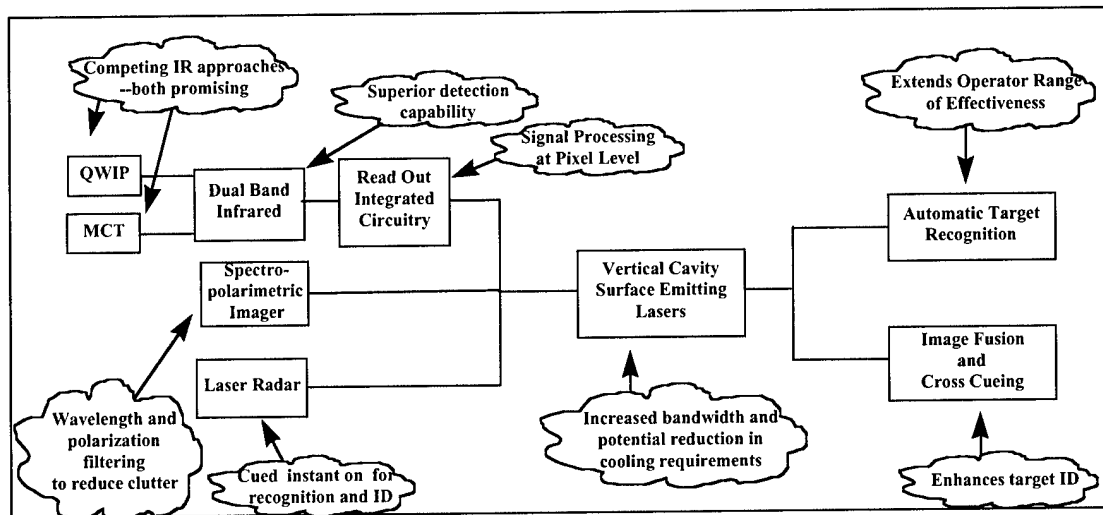


Figure 1. The Multi-Domain Smart Sensor combines several sensor types with advanced signal processing and aided target recognition for faster and more accurate battlefield threat classification.

The ultimate goal of the MDSS program is to demonstrate simultaneous active and passive infrared (IR) imaging over a wide spectral bandwidth. Another goal is to cue an active laser radar (LADAR) sensor using passive multi-color IR sensors to provide a dramatic improvement in battlefield situational awareness by rapid detection, location, and recognition of enemy targets (day or night, obscured and/or camouflaged) in highly cluttered environments. In the ultimate demonstrations, increased target detection range and reduced target classification time will be demonstrated using this advanced sensor hardware coupled with software developments (such as real-time sensor fusion hardware) from the signal processing and ATR technical factors.

The signatures of targets and backgrounds can vary significantly over the entire range of the IR spectrum. A sensor that can image simultaneously in different bands of the IR will have an advantage in target discrimination and clutter rejection over conventional single-band imaging systems. The active portion of the MDSS system is a laser radar system that will give a three-dimensional image of the target.

In the notional system, imagery from the (passive) multi-wavelength infrared sensors is processed to cue a LADAR which actively scans regions of interest to acquire high-resolution shape and range information for accurate and timely target classification using a combination of model-based and phenomenological ATR algorithms. In addition,

spectro-polarimetric imagery may be used to search for and match to specific non-imaging target features (such as chemical signatures) to cue an active sensor such as a LADAR. High-speed optical data paths using vertical-cavity surface-emitting lasers will provide thermal isolation and critical interconnect bandwidth for image transmission, processing, and sensor feedback.

A key element of the MDSS system is the dual-band IR imager (also known as a forward-looking infrared or FLIR). In the notional MDSS system, this imager consists of large-format, pixel-registered two-dimensional focal plane arrays (FPA) one of which is sensitive in the 3 to 5 μm mid-wave IR wavelength band (MWIR) and the other sensitive in the 8 to 12 μm long-wave IR wavelength band or (LWIR). Thus the passive part of the MDSS imager can take advantage of both of the atmospheric transmission bands in the IR spectrum.

Two approaches have been put forward to produce the dual-band IR imager portion of the MDSS. The first, being developed by DRS Infrared Technologies, Inc., uses the incumbent HgCdTe technology. This approach offers the advantage of near-unity quantum efficiency and an operating temperature near that of liquid nitrogen (77 K). The second approach, being employed by BAE Systems North America, uses quantum well IR photodetectors (QWIPs). The advantage of this approach is that the mature growth and processing technology of III-V compounds such as GaAs, AlGaAs, and InGaAs allow for greater array uniformity and higher yield relative to that of II-VI materials like HgCdTe. The disadvantage of QWIPs is that they have lower quantum efficiency relative to HgCdTe photodiodes and detectors operating in the LWIR spectral region need to be cooled to temperatures below 77 K (typically between 60 K and 65 K) to give background-limited performance (BLIP). Nevertheless, QWIPs have made great strides in recent years and now present a serious alternative to HgCdTe for high-performance IR imaging systems.

Experiment

The ultimate goal of the MDSS effort is that the dual-band FPA be 640 by 480 pixels or larger in both bands. However, the initial dual band FLIR format is to be 320 by 240 for DRS HgCdTe array and 256 by 256 for BAE QWIP FPA. The dual-band FLIR arrays were under development during 1998. Delivery of the dual-band arrays is expected during the second quarter of 1999.

From July 27 to 30 and September 14 to 18, 1998 field tests were held at the Drop Zone at Ft. A. P. Hill Military Reservation near Fredericksburg, VA. The goal of these field tests was to gather simultaneous IR imagery in the MWIR and LWIR bands of various military targets. Since the dual band FLIR was not available, separate MWIR and LWIR cameras were used for image acquisition. The cameras were configured such that the instantaneous fields-of-view (IFOV) of the pixels and the total fields of view of the cameras were the same in both the LWIR and MWIR bands. This was accomplished by choosing FPAs with equal pixel sizes and array formats as well as imaging lenses with

equal focal lengths. The properties of the cameras used are shown in table 1. The camera/data acquisition system is shown schematically in figure 2.

Table 1. Infrared cameras used in 1998 MDSS field tests

Property	MWIR	LWIR
FPA Manufacturer	Lockheed Martin Santa Barbara Focalplane	Sanders, a Lockheed Martin Company
Material Technology	InSb Photodiode	QWIP Photoconductor
Wavelength Range	3.0 to 5.3 μm	8.0 to 9.5 μm
Pixel pitch	24 μm by 24 μm	24 μm by 24 μm
Array format	640 by 480	640 by 480
Lens focal length	100 mm and 400 mm	100 mm and 400 mm
Focal ratio	f/2.5	f/2.3
IFOV	0.24 mrad and 0.06 mrad	0.24 mrad and 0.06 mrad
Total FOV	8.8° by 6.6° and 2.2° by 1.65°	8.8° by 6.6° and 2.2° by 1.65°
Operating temperature	77 K	62 K
Integration time	0.95 ms	1.83 ms
Temporal NEDT	0.025 K	0.032 K
Pixel operability*	99.84 %	99.25 %

*Defined as pixels with responsivity within $\pm 50\%$ from the mean

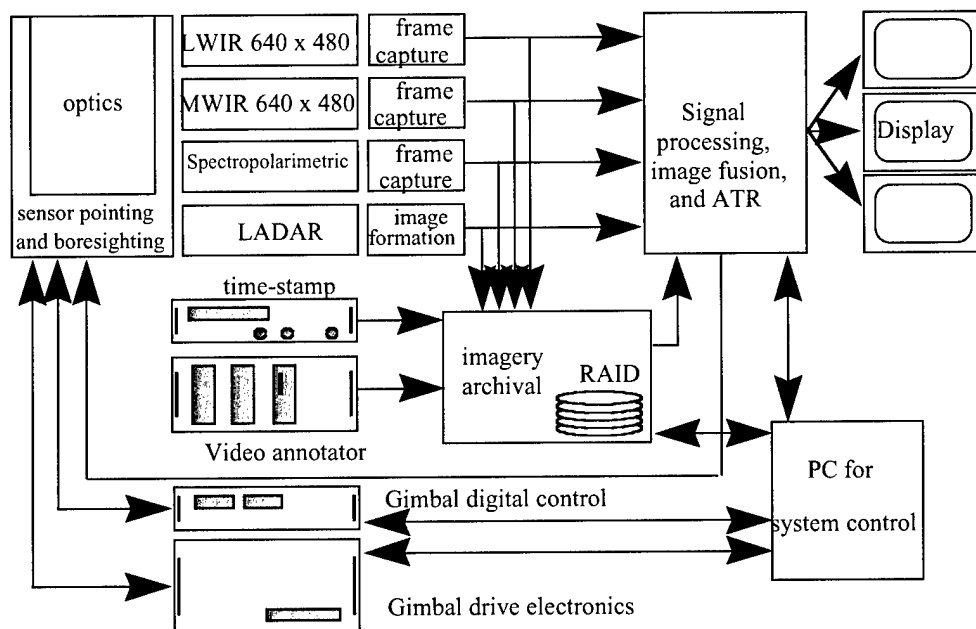


Figure 2. Schematic diagram of the camera/image acquisition system for MDSS field test.

A photograph of the cameras mounted on a computer-controlled gimbal is shown in figure 3. The gimbal had a pointing accuracy of 0.01° . The cameras were boresighted using micrometer-controlled optical mounts. A bright object that took up only a few pixels was identified in the LWIR image, and the position of the MWIR camera was adjusted such that the same object occupied the same pixel positions in the MWIR image. This method achieved perfect pixel registration. However, in practice the motion of the various components caused the images to be misregistered by 2 to 4 pixels in the horizontal and vertical directions in the narrow FOV (field of view) mode. Perfect registration (to within 1 pixel) was achieved for the wide FOV. We were able to confirm that the FOVs for each of the cameras were indeed the same in both wide and narrow field modes.

Several targets of military significance were imaged. Imagery was taken over a wide variety of ambient conditions during both day and night including scenarios just before and after sunrise and sunset. A list of the targets observed and their ranges is given in table 2. Over 1200 images were obtained for each of the MWIR and LWIR advanced FLIRs. Ground truth for MDSS-controlled vehicle tests include global positioning system target tracks and meteorological data. Three planned scenarios included:

Mock turntable scenarios: Each target vehicle was rotated (to driver-estimated accuracy) at 22° intervals to provide full rotation views at a fixed target elevation (ground-to-ground level elevation).

Smoke obscuration drills: In the first test, a stationary M60 tank at a range of 2100 m was obscured by hexachloroethane (HC) practice smoke. In the second test, the HC smoke obscured two stationary vehicles at fixed ranges (M2 Bradley at 3209 m and M113 APC at 1192 m).

Clutter/foilage obscuration drills: Each target vehicle was randomly driven over a clutter/occlusion course at a fixed range (1600 m).

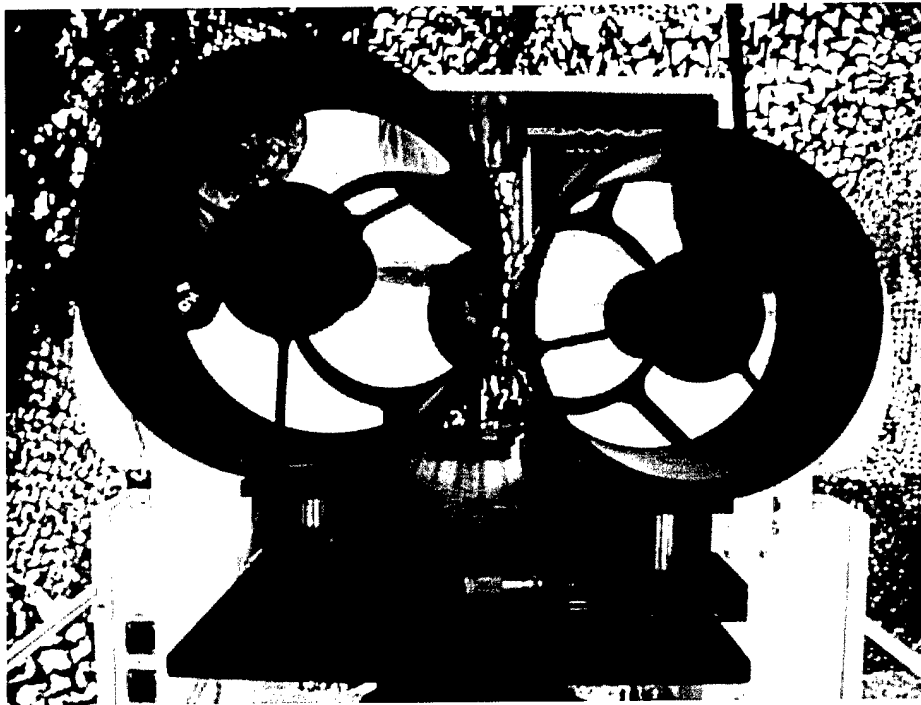


Figure 3. Boresighted IR cameras used in the MDSS field tests. The LWIR camera is on the left and the MWIR camera is on the right.

Table 2. Target vehicles and their ranges from the camera position

Vehicle	Ranges (m)
M60 Tank	1192, 3209
M2 Bradley Fighting Vehicle	1192, 2113, 3209, 4157
M35 Truck	1192, 3209
M113 Armored Personnel Carrier (APC)	1192, 2113, 3209, 4157
HMMWV	1192, 2113, 3209, 4157

An additional 75 images were obtained from sensors operated against targets of opportunity during Night Vision Electronic Sensors Directorate tests designed to measure the range and tracking capability of second-generation FLIRs operated on a YUHB-60 helicopter. Several panoramas, consisting of a series of 16 overlapping images, were taken of the Drop Zone at various times of day.

Smoke obscurants were tested during daylight operation to determine the visibility of stationary military targets at various ranges through HC smoke using the LWIR and MWIR imagers. Practice smoke from a K866 smoke pot was the obscurant. During the September field test, the stationary targets consisted of one M113 armored personnel carrier at 1192 meters range, and one M2 Bradley Fighting Vehicle at 3209 meters range; both vehicles were configured with standard Northern Forest camouflage paint. In the July field test, the stationary target was a single M60 tank at a range of 2100 meters.

Results and Discussion

a. Smoke

All imagery taken through HC smoke demonstrated greater target visibility with the LWIR camera than with either the MWIR or the visible light imagers. Typical results of the test are shown in figures 4 and 5. The visible light image shows complete obscuration of the M2 Bradley at 3209 m, and the partial obscuration of the M113 APC at 1192 m. The MWIR image shows partial obscuration of the M113 APC, although the M2 is visible. The LWIR provides ATR-quality imagery for both targets regardless of the high levels of HC smoke. The recent results confirm similar observations during earlier tests. Results from both tests showed that the LWIR QWIP camera imaged salient features of military vehicles which were obscured in the MWIR InSb and visible CCD camera imagery at ranges as far as 4157 meters.



Figure 4. Effect of HC smoke on imagery in the visible (a), MWIR (b), and LWIR (c). The vehicle in the foreground is an M113 APC at a range of 1192 m and that in the background is an M2 Bradley at 3209 m.

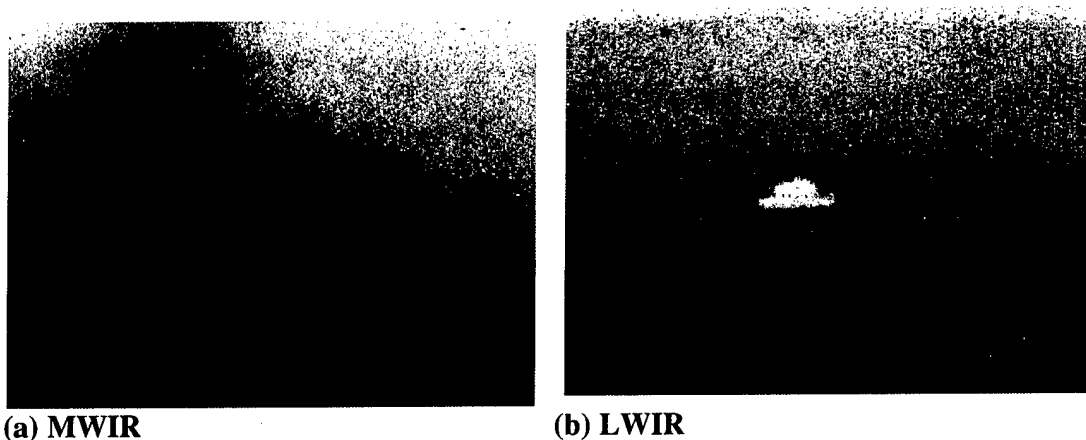


Figure 5. Effect of HC smoke on imagery in the MWIR (a) and LWIR (b). The target vehicle was an M60 tank at a range of 2100 m.

b. Ambient Conditions: Ground Fog and Rain

In the pre-dawn hours of September 15, the test range was shrouded in a heavy ground fog. The meteorological conditions at 0600 were as follows: temperature: 20.3 °C, relative humidity: 98%, and visibility: 1.8 km. Figure 6 shows imagery taken under these ambient conditions. The fog did not impact the MWIR imagery much. One can clearly see the tree line out to the end of the range, and cloud detail is visible in the sky. On the other hand, the LWIR image was severely degraded by the fog with the tree line invisible beyond about 2 km. It is interesting to note that the gravel road on the right side of the image appears to be bright in the MWIR image and dark in the LWIR image.

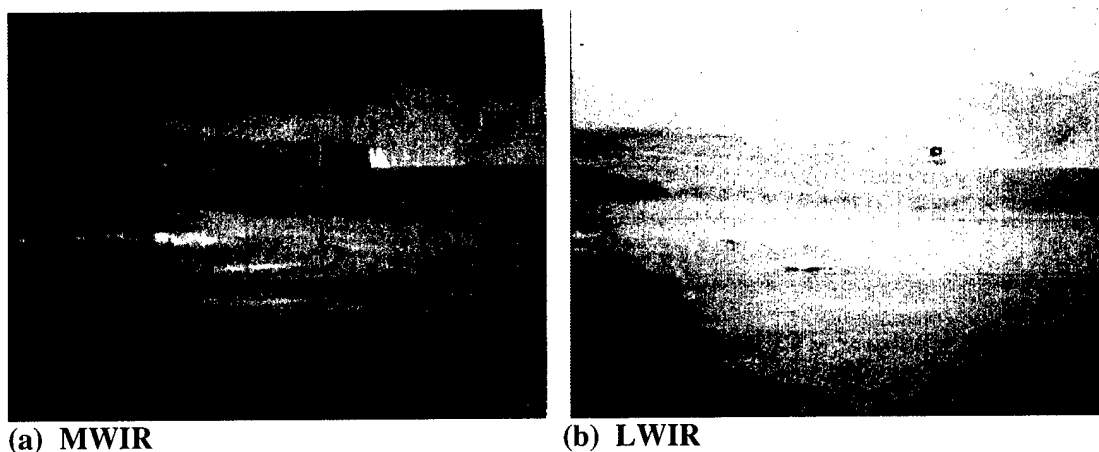


Figure 6. Images of the Ft. A. P. Hill dropzone taken before sunrise under conditions of heavy ground fog. The MWIR image (a) shows much detail downrange. The ground fog seriously degrades the LWIR image (b).

On the evening of September 17, heavy thunderstorms came through the area and caused the field test to be suspended. After the severe weather passed, the test resumed amid a light, steady rain (rain rate of approximately 1 mm/hr). The storms had cooled both the air and ground considerably: The air temperature dropped from 25.9 °C just before the storm to 20.3 °C after it had passed; the soil temperature dropped from 27.7 °C to 23.9 °C in the same period. The relative humidity after the storm was at or near 100% and the visibility was between 2 and 4 km.

Figure 7 shows MWIR and LWIR images taken just after the thunderstorm of an M113 APC (armored personnel carrier) and an M2 Bradley Fighting Vehicle at a range of 2 km. Figure 8 shows MWIR and LWIR images of the Bradley at a range of 4 km. The images shown in figures 7 and 8 consist of the central 320 by 240 pixels of the original 640 by 480-pixel images. The presence of rain and cooler air and ground temperatures caused both the LWIR and the MWIR image quality to be severely degraded. The MWIR imagery was affected by the ambient conditions to a greater extent than that of the LWIR. The M2 was clearly recognizable in both the LWIR and MWIR images at the 2 km range. However, the M113 was barely visible at all in the MWIR 2 km image. At the 4 km range, the M2 is almost lost in the noise of the MWIR image while the LWIR image still shows some detail of the vehicle as well as that of the tree line behind it.

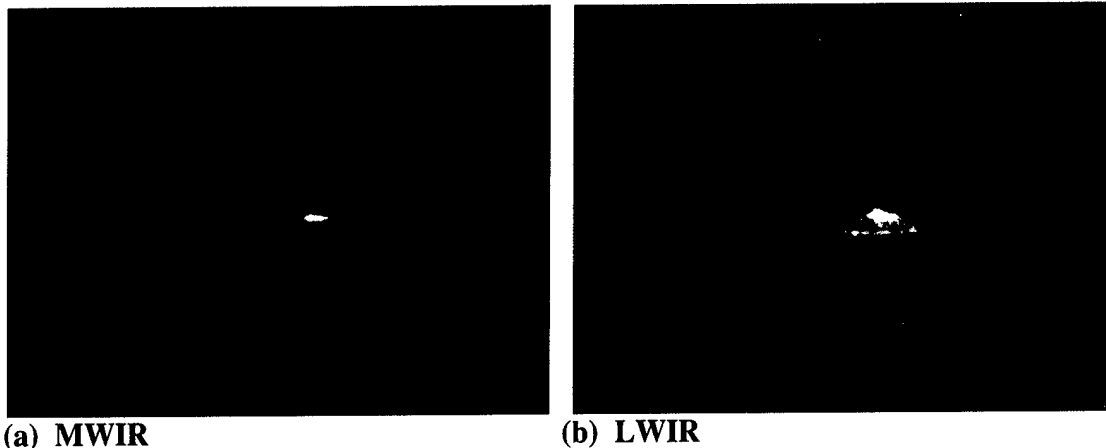


Figure 7. MWIR (a) and LWIR (b) Images of an M113 (left) and an M2 (right) taken at night under rainy conditions; the range to the targets was 2.1 km. The Figure shows the central 320 by 240 pixels of the acquired images. The M113 had been idle for approximately 2 h prior to these images while the M2 was running.

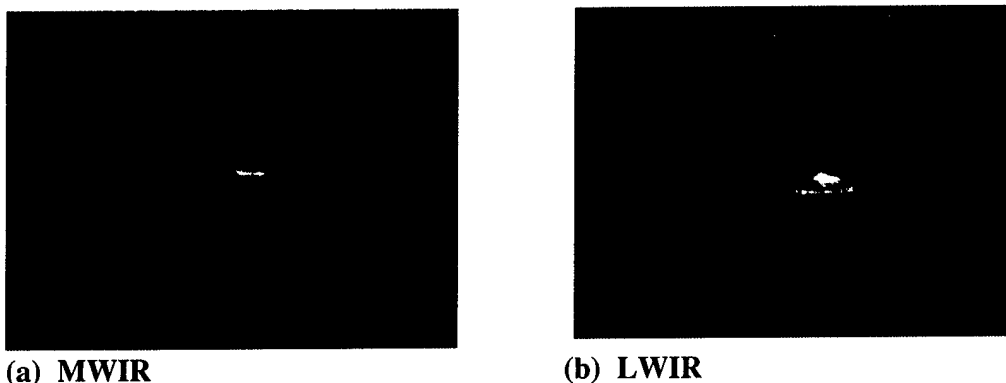


Figure 8. MWIR (a) and LWIR (b) Images of an M2 (right) taken at night under rainy conditions. The range to the targets was 4 km. The Figure shows the central 320 by 240 pixels of the acquired images.

c. Image Fusion

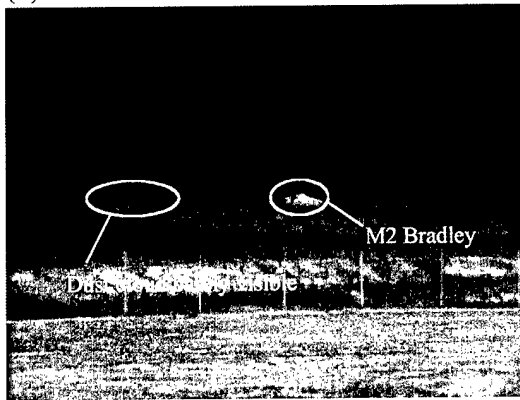
The goal of dual-band or multicolor IR imagery is to provide more information about the target and/or background to a human observer or to an automatic target recognition system than could be provided by a single band imager. To present this additional information to the user, the dual-band image data needs to be combined (fused) into a single image. Many methods have been proposed to do the fusion, but the most straightforward methods of image fusion are the simple sum and difference of the individual images.

For all image fusion methods it is important that the individual images be equalized with each other. The pixel values in each image ranged from 0 to 4095 (12 bits). The majority of the pixel values covered a spread of approximately 200 counts near the center of the range. The differences between the pixel values and the value at which the peak of the histogram occurred were calculated. The resulting pixel values formed the equalized images for both the LWIR and the MWIR images. The equalized images could then be summed or subtracted from one another to give new images that combined the

information in both of the individual images. An example of the results of this process is shown in figure 9. The MWIR and LWIR data are shown as grayscale images with hot objects represented as white and cold objects as black. The hot engine exhaust on the side of the vehicle (an M2 Bradley Fighting Vehicle) shows up bright in both images. The exposed dirt just beyond the fence in the foreground is hotter than the surrounding grass-covered ground, which is, in turn, warmer than the trees in the background. In the MWIR image there is a region behind the vehicle that is slightly brighter than its surroundings.

Figure 9 (c) shows the result of subtracting the pixel values of the MWIR image from those of the LWIR image. In the difference image, white pixels indicate regions where the LWIR intensity is greater than that of the MWIR, while dark pixels are those regions where the MWIR intensity dominates that of the LWIR. In the fused difference image, the entire dust plume kicked up by the moving vehicle is visible. It is only through the fusion of the two single-color images that the dust plume becomes plainly visible.

(a) MWIR



(b) LWIR



(c) Fused Difference LWIR—MWIR

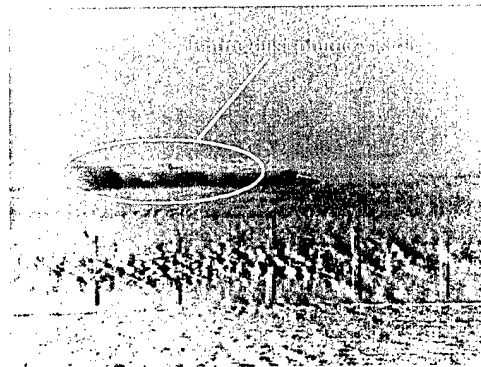


Figure 9. IR images of a Bradley Fighting Vehicle (M2) in the MWIR (a) and LWIR (b). The vehicle was moving from left to right across the frame at a range of approximately 500 meters. The fused difference image is shown in (c).

The remaining examples of image fusion are the result of a more sophisticated color fusion algorithm, developed at the Naval Research Laboratory [ii] (NRL), in which pixel values from the LWIR and MWIR bands are assigned to color opponents such as red-cyan, blue-yellow, or green-magenta. Figure 10 illustrates this color fusion scheme using the red-cyan color opponents. Each pixel in the LWIR image is assigned a red value and each pixel in the MWIR image is assigned a cyan value (i.e., equal values of blue and green). For 8-bit color, the pixel values range from 0 to 255. Objects in the image with high brightness values in both bands will appear white; those with low brightness values in both bands will appear black. Objects with a high pixel value in the LWIR band and a low value in the MWIR band will appear red, and objects with a low pixel value in the LWIR band and a high value in the MWIR band will appear cyan.

In this scheme, bands in which the background and targets are highly correlated will yield fused images with little color contrast (the pixel data will lie roughly along the diagonal, $[0,0]$ to $[255,255]$, of the plot and the image will appear as shades of gray). Bands in which they are weakly correlated will yield maximum color contrast (the pixel data will be spread out in a direction orthogonal to the diagonal). In the case where the background is highly correlated and the target is only slightly different, the color contrast can be enhanced by performing a principal component (PC) transformation, normalizing the data along the PC directions (thereby stretching the data to fill the available color space), then performing the inverse transform back to the original color-opponent space.

An example of this color fusion approach is shown in figure 11. The LWIR and MWIR images were taken in the wide FOV configuration in early morning near dawn with a significant amount of ground fog present (the individual MWIR and LWIR images are shown in fig. 6). The tree line is between 1 and 3 kilometers from the cameras. Image fusion using blue-yellow color opponents, in which the pixel values of the MWIR image are mapped to shades of yellow and those of the LWIR image are mapped to shades of blue, yields an image with the sky looking blue and the grass looking green giving a realistic visual feel while still conveying the thermal characteristics of the scene.

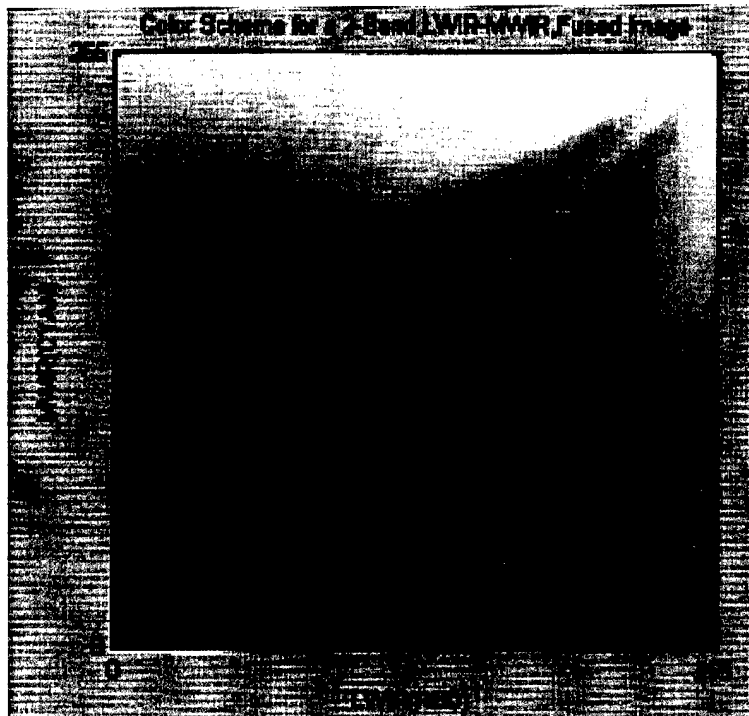


Figure 10. Red-cyan color scheme for fusing MWIR and LWIR images. The MWIR image is mapped to shades of cyan and the LWIR image is mapped to shades of red. Pixels with approximately equal values in the MWIR and LWIR will lie along the diagonal of the color diagram and will be represented by shades of gray.

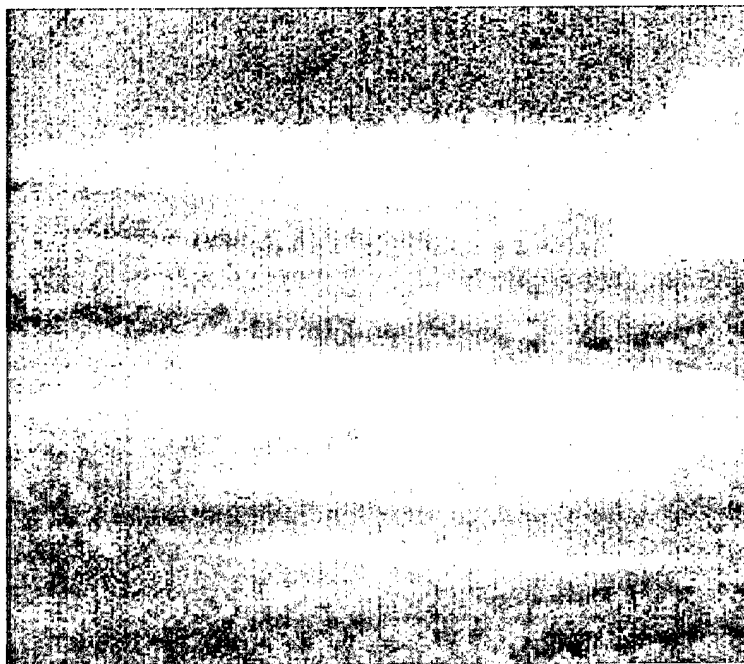


Figure 11. Color fusion algorithm applied to the MWIR and LWIR images shown in fig. 6. The MWIR pixel values are mapped to shades of yellow and the LWIR values to shades of blue. The color fusion gives a realistic feel to the image (blue sky and green grass).

An advantage of this color fusion approach over monochrome fusion approaches, such as the difference image discussed above, is that it not only displays information about which objects are bright, but it also displays information about the band in which the object is emitting. This is demonstrated in figure 12 which shows, respectively, the fused difference (a) and color-fused (b) images of an M60 tank taken through HC smoke. The corresponding MWIR and LWIR images were shown previously in figure 5. The tank is visible only in the LWIR image and therefore is a bright red in the color-fused image. The grass, trees, and sky that are difficult to distinguish in the difference image are clearly separated in the color-fused image.

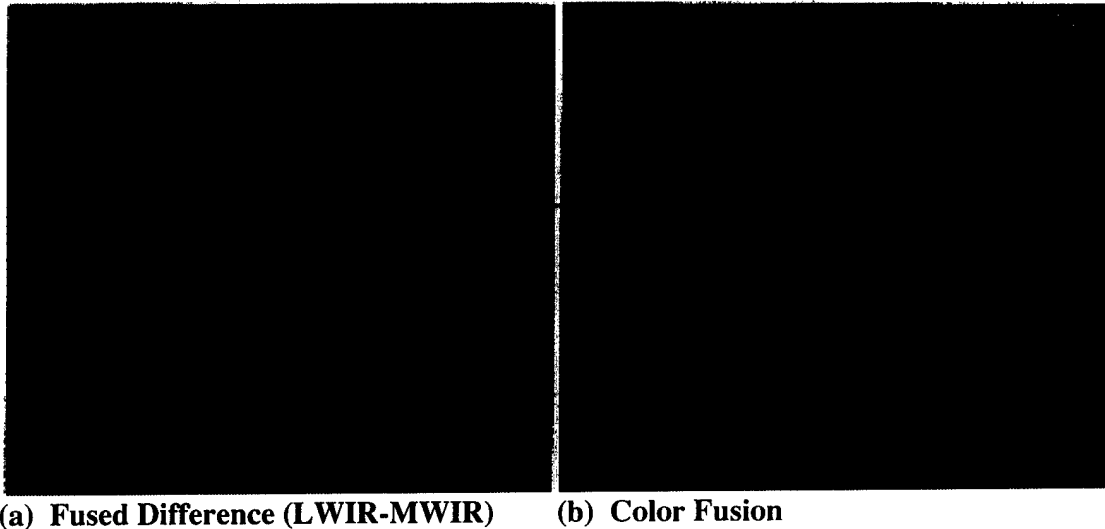


Figure 12. Fused IR imagery of an M60 tank through HC smoke (individual MWIR and LWIR images shown in fig. 5). Monochrome difference image is shown in (a). Color fusion with LWIR mapped to shades of red and MWIR to shades of cyan is shown in (b). The color fusion clearly shows more detail of the scene including the smoke, the background and foreground vegetation as well as the tank.

d. ATR

As the quantity of information that helicopter and tank crews must analyze increases, the need for an automated screening process increases. It is anticipated that vehicle crews will not be able to operate the vehicle and simultaneously view the outputs of MWIR, LWIR, visible, and LADAR sensors. An algorithm that screens the data and presents only the most likely targets to the operator would enable the crew to make maximum use of the data.

To test the utility of dual band IR imagery, automated target detection and clutter rejection (CR) algorithms were designed, coded, and run on the MDSS data collected at Ft. A. P. Hill. The idea behind the experiments was to quantify algorithm performance on the MDSS data set using MWIR data only, LWIR data only, and MWIR and LWIR together. If an algorithm performs better on both bands together, then there is some utility in having a dual band sensor. If not, then this suggests that the data is nearly redundant, that almost all of the information in one band is contained in the other.

The experiment was performed by applying a simple detector to each image separately, and counting as a detection any location that was reported by the detector on either of the images (i.e., the detection locations for both bands is a superset of the detection locations for each band). Image chips were formed by extracting a target size region from the image at each detection location and scaling to a standard range, so that each chip is the size of a target size region at the standard range. This allows the use of a learning algorithm that is not scale invariant. The image chips were then separated into disjoint training and testing sets. The chips were used as input to three clutter rejectors: one operating on MWIR alone, one on LWIR alone, and MWIR-LWIR together.

The detection algorithms were simple untrained algorithms that look for regions of approximately the size of the target that display some difference from their immediate background. A detailed description can be found in Dwan and Der [iii]. The mathematical features that were used to determine if a difference existed include gray level (hot or cold spots), local variance, component size blobs, edge strength, and so on. The features were combined with a weighted sums algorithm. Since the algorithm is nearly untrained, it should, and does, perform about equally well on MWIR and LWIR. Figure 13 shows the detection rates on the training and test sets, as a function of false alarms per frame.

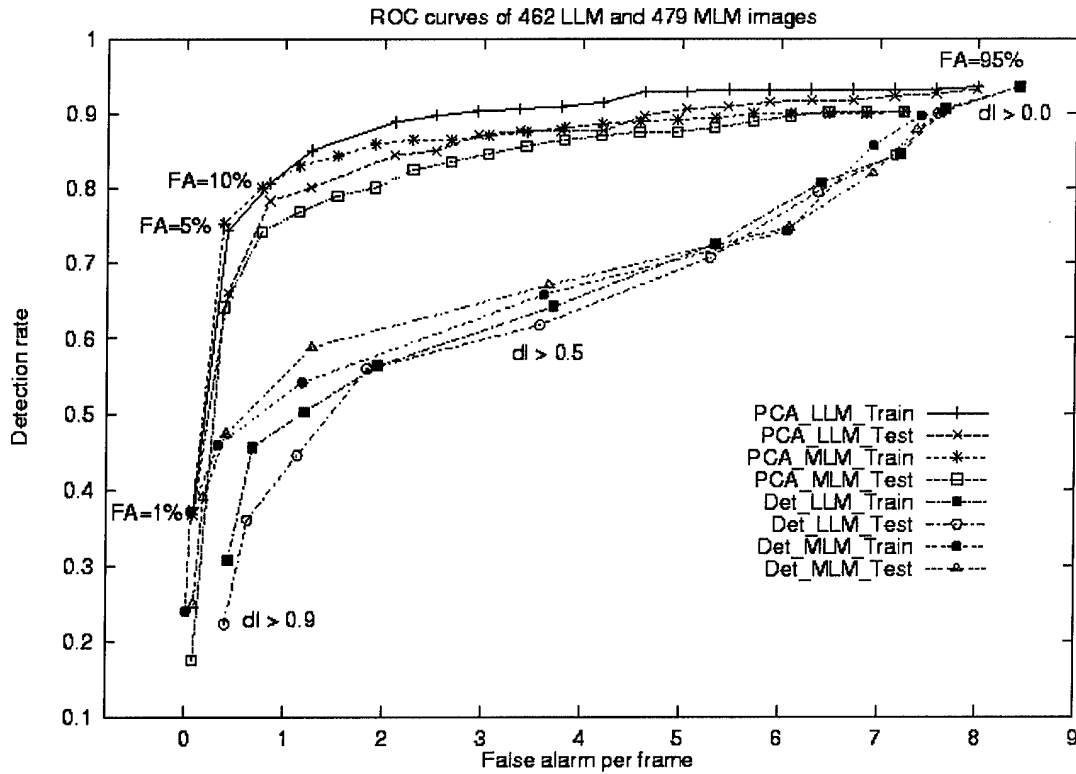


Figure 13. Detection rates on the training and test sets, as a function of false alarms per frame.

The clutter rejection algorithms used in the experiments were based on principal component analysis (PCA) or eigenspace separation transform (EST) reduction of the data, followed by a neural network. The PCA/EST portion of the algorithm was applied to the training set to compress the imagery into the few parameters that describe the most of the variability in the set of images. The compression was then applied to the test imagery, and the resulting components were input to a neural network which had been trained to distinguish between clutter and target components. For the case which used both MWIR and LWIR data together, the image vectors were simply appended. Description of the PCA and EST transforms are given below, followed by a description of the neural network that uses these features.

1. PCA

PCA, also referred to as the Hotelling transform or the discrete Karhunen-Loève transform, is based on statisticaures. PCA is an important tool for image processing because it has several useful properties, such as decorrelation of data and compaction of information (energy) [iv]. We provide here a summary of the basic theory of PCA.

Assume a population of random vectors of the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (1)$$

The *mean vector* and the *covariance matrix* of the vector population \mathbf{x} are defined as

$$\mathbf{m}_x = E\{\mathbf{x}\}, \quad \text{and} \quad (2)$$

$$\mathbf{C}_x = E\{(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^T\}, \quad (3)$$

where $E\{\arg\}$ is the expected value of the argument, and T indicates vector transposition. Because \mathbf{x} is n -dimensional, \mathbf{C}_x is a matrix of order n by n . Element c_{ii} of \mathbf{C}_x is the variance of x_i (the i th component of the \mathbf{x} vectors in the population), and element c_{ij} of \mathbf{C}_x is the covariance between elements x_i and x_j of these vectors. The matrix \mathbf{C}_x is real and symmetric. If elements x_i and x_j are uncorrelated, their covariance is zero and, therefore, $c_{ij} = c_{ji} = 0$. For N vector samples from a random population, the mean vector and covariance matrix can be approximated from the samples by

$$\mathbf{m}_x = \frac{1}{N} \sum_{p=1}^N \mathbf{x}_p, \quad \text{and} \quad (4)$$

$$\mathbf{C}_x = \frac{1}{N} \sum_{p=1}^N (\mathbf{x}_p \mathbf{x}_p^T - \mathbf{m}_x \mathbf{m}_x^T) \quad (5)$$

Because \mathbf{C}_x is real and symmetric, we can always find a set of n orthonormal eigenvectors for this covariance matrix. A simple but foolproof algorithm to find these orthonormal eigenvectors for all real symmetric matrices is the Jacobi method [v]. The Jacobi algorithm consists of a sequence of orthogonal similarity transformations. Each transformation is just a plane rotation designed to annihilate one of the off-diagonal matrix elements. Successive transformations undo previously set zeros, but the off-diagonal elements get smaller and smaller, until the matrix is effectively diagonal (to the precision of the computer). We obtain the eigenvectors by accumulating the product of transformations during the process, while the main diagonal elements of the final diagonal matrix are the eigenvalues. Alternatively, a more complicated method based on the QR algorithm for real Hessenberg matrices can be used [vi]. This is a more general method because it can extract eigenvectors from a nonsymmetric real matrix. Furthermore, it becomes increasingly more efficient than the Jacobi method as the size of the matrix increases. Given the considerable increase in efficiency for the size of our covariance matrix, we chose the QR method for our experiments described in this paper. Figure 14 shows the first 50 most dominant PCA eigenvectors representing the targets (top 5 rows) and clutter (bottom 5 rows) in the training set. Having the largest eigenvalues, these eigenvectors capture the greatest variance or energy as well as the most meaningful features among the training data.

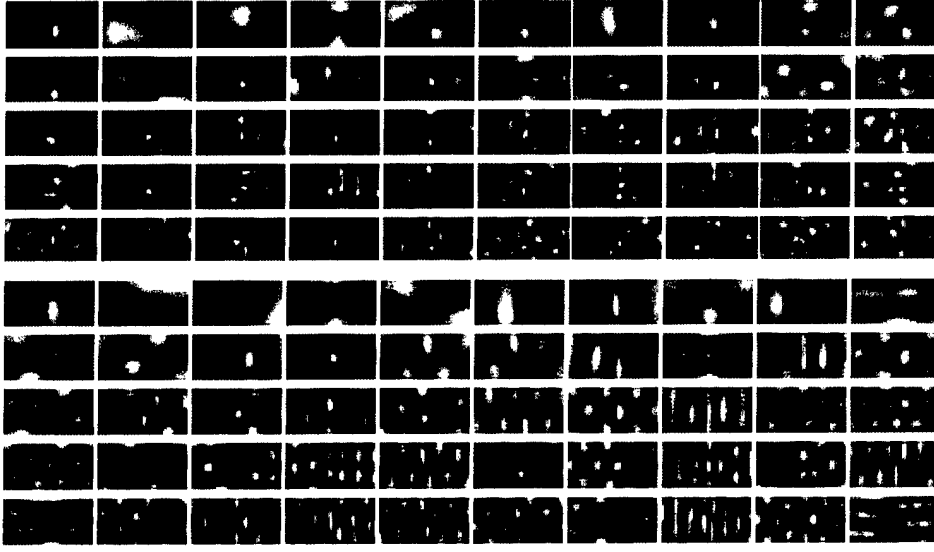


Figure 14. First 50 most dominant PCA eigenvectors for the targets (top 5 rows) and clutter (bottom 5 rows) in the training set.

Let \mathbf{e}_i and λ_i , $i = 1, 2, \dots, n$, be the eigenvectors and the corresponding eigenvalues of \mathbf{C}_x , sorted in a descending order so that $\lambda_j \geq \lambda_{j+1}$ for $j = 1, 2, \dots, n - 1$. Let \mathbf{A} be a matrix whose rows are formed from the eigenvectors of \mathbf{C}_x , such that

$$\mathbf{A} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}. \quad (6)$$

This \mathbf{A} matrix can be used as a transformation matrix that maps the \mathbf{x} 's into vectors denoted by \mathbf{y} 's, as follows:

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_x). \quad (7)$$

The \mathbf{y} vectors resulting from this transformation have a zero mean vector; that is, $\mathbf{m}_y = \mathbf{0}$. The covariance matrix of the \mathbf{y} 's can be computed from \mathbf{A} and \mathbf{C}_x by

$$\mathbf{C}_y = \mathbf{A} \mathbf{C}_x \mathbf{A}^T. \quad (8)$$

Furthermore, \mathbf{C}_y is a diagonal matrix whose elements along the main diagonal are the eigenvalues of \mathbf{C}_x ; that is,

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}. \quad (9)$$

Since the off-diagonal elements of \mathbf{C}_y are zero, the elements of the \mathbf{y} vectors are uncorrelated. Since the elements along the main diagonal of a diagonal matrix are its eigenvalues, \mathbf{C}_x and \mathbf{C}_y have the same eigenvalues and eigenvectors. In fact, the transformation of the \mathbf{C}_x into \mathbf{C}_y is the essence of the Jacobi algorithm described above.

Therefore, through the PCA transformation, a new coordinate system is established. The origin of this new coordinate system is at the centroid of the population, \mathbf{m}_x , with new axes in the direction specified by the eigenvectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$. The eigenvalue λ_i becomes the variance of component y_i along eigenvector \mathbf{e}_i . With its ability to realign unknown data into a new coordinate system based on the principal axes of the data, PCA is often used to achieve rotational invariance in image processing tasks.

On the other hand, we may want to reconstruct vector \mathbf{x} from vector \mathbf{y} . Because the rows of \mathbf{A} are orthonormal vectors, $\mathbf{A}^{-1} = \mathbf{A}^T$. Therefore, any vector \mathbf{x} can be reconstructed from its corresponding \mathbf{y} by the relation

$$\mathbf{x} = \mathbf{A}^T \mathbf{y} + \mathbf{m}_x. \quad (10)$$

Instead of using all the eigenvectors of \mathbf{C}_x , we may pick only k eigenvectors corresponding to the k largest eigenvalues and form a new transformation matrix \mathbf{A}_k of order $k \times n$. In this case, the resulting \mathbf{y} vectors would be k -dimensional, and the reconstruction given in eq. (10) would no longer be exact. The reconstructed vector using \mathbf{A}_k is

$$\hat{\mathbf{x}} = \mathbf{A}_k^T \mathbf{y} + \mathbf{m}_x. \quad (11)$$

The mean square error (MSE) between \mathbf{x} and $\hat{\mathbf{x}}$ can be computed by the expression

$$\varepsilon = \sum_{j=1}^n \lambda_j - \sum_{j=1}^k \lambda_j = \sum_{j=k+1}^n \lambda_j. \quad (12)$$

Because of the λ_j 's decrease monotonically, eq. (12) shows that we can minimize the error by selecting the k largest eigenvalues. Thus, the PCA transformation is optimal in the sense that it minimizes the MSE between the vectors \mathbf{x} and their approximations $\hat{\mathbf{x}}$.

The EST has been proposed by Torrieri as a preprocessor to a neural binary [vii]. The goal of the EST is to transform the input patterns into a set of projection values such that the size of a neural classifier is reduced and its generalization capability is increased. The size of the neural network is reduced, because the EST projects an input pattern into an orthogonal subspace of smaller dimensionality. The EST also tends to produce projections with different average lengths for different classes of input and, hence, improves the discriminability between the targets. In short, the EST preserves and enhances the classification information needed by the subsequent classifier. It has been used in a mine-detection task with some success [viii].

The transformation matrix \mathbf{S} of the EST can be obtained as follows.

Compute the n by n correlation difference matrix

$$\hat{\mathbf{M}} = \frac{1}{N_1} \sum_{p=1}^{N_1} \mathbf{x}_{1p} \mathbf{x}_{1p}^T - \frac{1}{N_2} \sum_{q=1}^{N_2} \mathbf{x}_{2q} \mathbf{x}_{2q}^T, \quad (13)$$

where N_1 and \mathbf{x}_{1p} are the number of patterns and the p th training pattern of Class 1, respectively. N_2 and \mathbf{x}_{2q} are similarly related to Class 2 (which is the complement of Class 1).

1. Calculate the eigenvalues of $\hat{\mathbf{M}}$, $\{\lambda_i \mid i = 1, 2, \dots, n\}$.
2. Calculate the sum of the positive eigenvalues

$$E_+ = \sum_{i=1}^n \lambda_i \quad \text{if } \lambda_i > 0, \quad (14)$$

and the sum of the absolute values of the negative eigenvalues

$$E_- = \sum_{i=1}^n |\lambda_i| \quad \text{if } \lambda_i < 0. \quad (15)$$

- (a) If $E_+ > E_-$, then take all the k eigenvectors of $\hat{\mathbf{M}}$ that have positive eigenvalues and form the n by k matrix \mathbf{S} .
- (b) If $E_+ < E_-$, then take all the k eigenvectors of $\hat{\mathbf{M}}$ that have negative eigenvalues and form the n by k matrix \mathbf{S} .
- (c) If $E_+ = E_-$, then use either subset of eigenvectors to form the matrix \mathbf{S} , preferably the smaller subset.

Given the \mathbf{S} transformation matrix, the projection \mathbf{y}_p of an input pattern \mathbf{x}_p is computed as $\mathbf{y}_p = \mathbf{S}^T \mathbf{x}_p$. The \mathbf{y}_p , with a smaller dimension (because $k \leq n$) and presumably larger separability between the classes, can then be sent to a neural classifier. Figure 15 shows the eigenvectors associated with the positive and negative eigenvalues of the $\hat{\mathbf{M}}$ matrix that was computed with the target chips as Class 1 and the clutter chips as Class 2. From the top 5 rows of the figure, we may trace those signatures that are associated with the targets. On the other hand, the bottom 5 rows represent mostly features of the clutter. As shown in figure 16, while the eigenvalues diminish rapidly for both the PCA and EST methods, those of the EST decrease even faster. In other words, the EST may produce a higher compaction in contextual information.

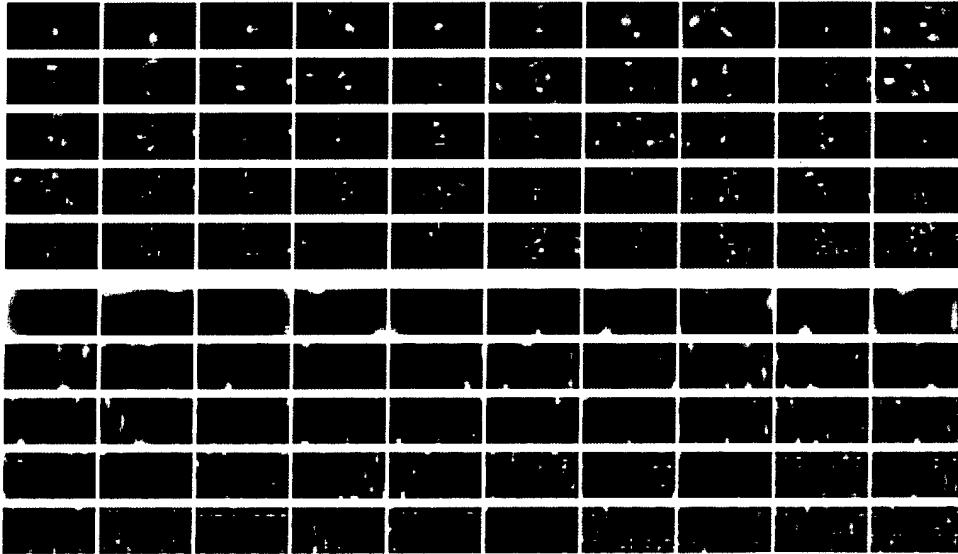


Figure 15. First 50 most dominant EST eigenvectors associated with positive (top 5 rows) and negative (bottom 5 rows) eigenvalues for the training set.

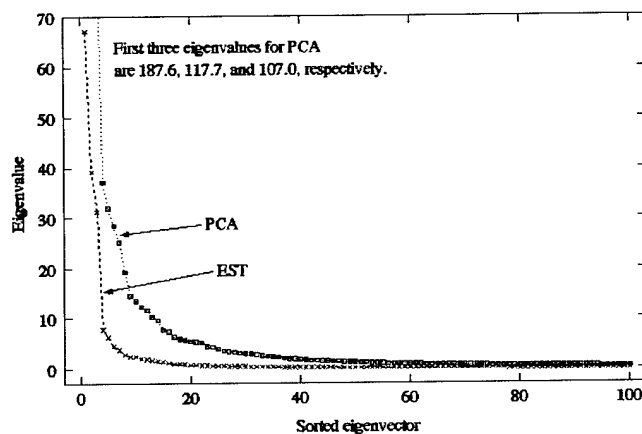


Figure 16. Rapid attenuation of eigenvalues in PCA and EST transforms.

2. Clutter Rejection

The inputs for our clutter rejection module are the image chips extracted from bigger scenes. The size of these image chips is fixed to a predefined dimension, which is common to both the targets and the clutter. To reduce the background information in target chips, we clip each image chip at a size that equals the dimension of the largest target in our training set. After the background removal, the input image is scaled to a preferred size based on a linear interpolation technique. This scaling is needed to achieve an image size that is efficient for feature extraction via the eigenspace transformation, while an effective amount of information is retained in the image.

After normalizing the clipped and scaled training data, we compute the eigenvectors using either PCA or the EST. We treat each image pixel as a dimension of the data vector in these computations. The resulting eigenvectors are sorted in descending order based on the norm of their corresponding eigenvalues. Characterized by their eigenvalues, different subsets of these eigenvectors may be used as feature extractors in different experiments. To achieve feature extraction and dimensionality reduction, we project the preprocessed input image to a chosen set of n eigenvectors. The resulting n projection values are fed to a multi-layer perception (MLP) algorithm, where they are nonlinearly combined.

A typical MLP used in our experiments is shown in figure 17. The MLP has $n+1$ input nodes (with an extra bias input), several layers of hidden nodes, and one output node. In addition to full connections between consecutive layers, there are also shortcut connections directly from one layer to all other layers, which may speed up the learning process. The MLP is trained to perform a two-class problem, with training output values of ± 1 . Its sole task is to decide whether a given input pattern is a target (indicated by a high output value of around $+1$) or clutter (indicated by a low output value of around -1). The MLP is trained in batch mode by a modified Qprop algorithm [v] for a quick but stable learning course.

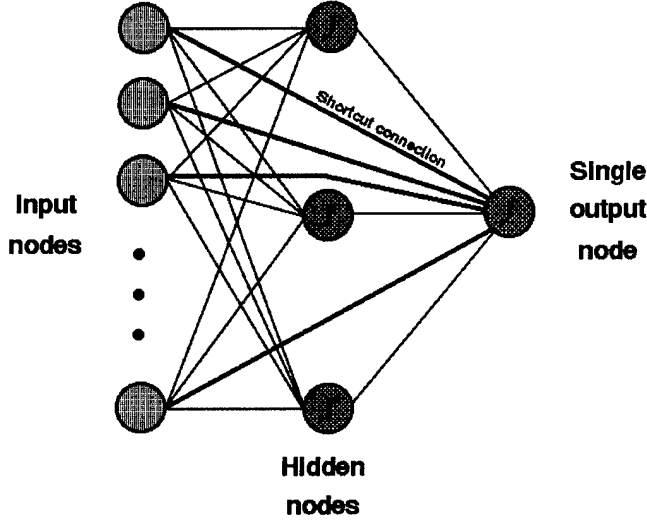


Figure 17. A simple MLP with two layers of weights and shortcut connections.

If the number of target chips and clutter chips are quite different in the training set, a trained MLP tends to predict the class that has more training samples. This negative effect of an *imbalanced* training set has been studied by Anand, *et al.*[ix]. To avoid creating such a biased network, we add a corrective measure in our modified learning algorithm. Because the training is carried out in batch mode [x], the *error gradient* $\frac{dE}{dw}$ obtained for each network parameter or *weight* for a given training pattern can be accumulated separately, depending on the type of intended outputs for that training pattern. At the end of a training epoch, the average value of the error gradient when the training output is high (low), ϵ^h , (ϵ^l), for a weight i is computed as

$$\epsilon_i^h = \frac{1}{N_h} \sum_{p=1}^{N_h} \frac{dE_p^h}{dw_i} \quad \text{and} \quad \epsilon_i^l = \frac{1}{N_l} \sum_{p=1}^{N_l} \frac{dE_p^l}{dw_i} \quad (16)$$

where N_h and N_l are the number of occurrences of high and low training objects, respectively. If ϵ_i^h and ϵ_i^l have the same sign or direction, then their average is used to update the corresponding weight i . Otherwise, no update is made to the controversial weight. This corrective scheme allows the output errors incurred by both high and low target outputs to be reduced simultaneously. To maximize the class separation between the targets and clutter, we focus only on the training patterns that are easily confused or wrongly classified at a predefined false-alarm rate. Only the errors incurred by these confusing patterns are used to update the MLP weights, so that these patterns may be classified correctly later. A less confusing pattern may be considered only during the early stage of training.

This technique of focused learning improves the target recognition rate drastically for a given false-alarm rate.

3. Experimental Results

To examine the performance of our clutter rejection technique, we implement a difficult two-class problem. Second-generation 10-bit gray-scale FLIR images of five target types were obtained at three separate sites during different seasons of the year. These region of interest images were purposely captured under challenging conditions, such as having targets in and around clutter, in different backgrounds, and under various weather conditions. We used a neural-based target detector (developed at ARL by Christopher Dwan and Sandor Der) to detect the potential target areas in these images. The detected areas were then extracted as image chips of size 75 by 40 pixels, and labeled as either a target or clutter based on the ground-truth information. Because the target locations suggested by the detector might not match well with the ground-truth locations, and no manual centering was performed during the extraction process, many silhouettes remain severely off-center in the resulting target chips. There were 47,716 image chips in our training set, in which 4,627 were target chips and 43,089 clutter chips. On the other hand, there were 2,459 target chips and 18,070 clutter chips in the testing set. The testing set and 29,053 chips of the training set were taken from the same site, but in a different month and year.

Considering the size of the targets and the computational complexity of the QR algorithm (which is roughly proportional to the cube of the image size), we scale the input image to a moderate size of 40 by 20 pixels. As shown in figure 16, the norms of the eigenvalues also decrease rapidly from their respective maximum values in both types of eigenspace transformation. Therefore, we were only interested in the 40 most dominant eigenvectors, instead of all 800 eigenvectors available.

For PCA, the covariance matrix is computed from all the target images in the training set. For EST, on the other hand, the target images in the training set form Class 1, while the clutter images form Class 2. We used the 1, 5, 10, 20, 30, and 40 most dominant eigenvectors of each transformation to produce the projection values for the MLP. In each case, five independent training processes were tried with different initial random weights for the MLP.

When the MLP has fewer than 40 inputs, the average recognition rates for both PCA and EST increase with the number of eigentargets used for feature extraction. With 40 or more inputs, however, their performances started to either saturate or drop, indicating that the larger MLPs might have become over-fitted to the training set. When fewer than 20 projection values are used, the EST performed better than PCA. This improvement can be attributed to the better compaction of information associated with EST. On the other hand, the slightly lower recognition rates achieved by the EST with 20 or more inputs indicate that some minor information might have been lost in this transformation. Because a smaller number of inputs implies a much simpler and faster MLP, it would be most suitable to use EST in situations where speed and efficiency are more of a concern than slightly degraded recognition performance. In other situations, PCA is more suitable for achieving the maximum recognition performance possible through a bigger and slower MLP.

For the two-band case, the CR was implemented in two ways. First, the input LWIR and MWIR chips were appended, to form one vector, which was used to train the PCA and

EST algorithms. The resulting outputs were applied to an MLP in the same manner as described above. The second method uses the previously trained PCA and EST basis functions in parallel, resulting in twice as many outputs (LWIR plus MWIR outputs). The outputs were then input to an MLP with twice as many input nodes.

The threshold on each CR was set to allow a false alarm (FA) rate of 10 percent. Table 3 gives a breakdown of the images and chips available for the study. Tables 4 and 5 show the performance of the PCA CR and EST CR on the LWIR data only. Likewise tables 6 and 7 give performance on the MWIR data, tables 8 and 9 give performance for both bands together using the first method, and tables 10 and 11 using the second method. Note that the first multiband method gives slightly better performance than the second method. Also, in all cases, the maximum performance corresponds to 20 eigenvectors. In all cases, PCA gives maximum performance superior to EST. However, if the number of eigenvectors is fixed at a low level, the EST gives superior performance in some cases, implying that EST will be useful for applications that require low computational complexity.

The maximum target hit rates for the four CRs were 90.34, 87.34, 93.49 and 93.31 percent, for the MWIR, LWIR, and two multiband CRs, respectively. In other words, the multiband CR was able to reduce the missed detections by 51.42 percent for a fixed level of false alarms, over LWIR alone, and 32.6 percent over MWIR alone.

A word of caution is in order here about the relative merits of LWIR versus MWIR. While the results here suggest that MWIR is superior to LWIR, it is quite possible that the difference is due more to the particular sensors brought to the data collection than to the inherent physical limitations of the two bands. The opinion of the majority of the IR community is that, for state of the art sensors, LWIR gives superior quality to MWIR. Regardless of whether this is true, the experiments here suggest that the two bands are sufficiently independent of each other that multiband IR gives performance superior to a single band, as long as the single bands give similar performance alone.

Table 3. The number of training/testing image clips used for the clutter-rejection study.

Purpose	Data	Target	Clutter	Total
Training	LLM	273	1906	2179
	LBM	282	1861	2143
	MLM	282	1861	2143
	MBL	273	1906	2179
Testing	LLM	272	1906	2178
	LBM	281	1860	2141
	MLM	281	1860	2141
	MBL	272	1906	2178

Table 4. Hit rates of PCA-d2b_L (LLM chips detected by 2 bands: LLM+LBM) at 10 percent FA rate. The MLP has either 1, 5, 10, 20, 30, or 40 inputs plus a bias.

Number of inputs	Data type	Hit rates at 10 percent FA of five runs (%)					
		1	2	3	4	5	Avg.
1	Train	43.96	43.96	43.96	43.96	43.96	43.96
	Test	37.79	37.79	37.79	37.79	37.79	37.79
5	Train	84.50	84.32	84.14	86.31	85.23	84.90
	Test	74.50	78.30	76.31	77.94	76.85	76.78
10	Train	89.37	85.95	84.68	91.53	86.85	87.68
	Test	86.26	79.39	77.03	86.26	81.92	82.17
20	Train	92.97	91.17	94.95	91.89	96.04	93.40
	Test	86.80	86.44	88.43	84.99	90.05	87.34
30	Train	88.29	95.68	90.45	91.53	80.72	89.33
	Test	85.17	88.79	82.64	87.34	77.03	84.19
40	Train	84.68	88.11	83.06	82.70	82.52	84.21
	Test	80.83	86.08	78.12	78.12	78.30	80.29

Table 5. Hit rates of EST-d2b_L (LLM chips detected by 2 bands: LLM+LBM) at 10 percent FA rate. The MLP has either 1, 5, 10, 20, 30, or 40 inputs plus a bias.

Number of inputs	Data type	Hit rates at 10 percent FA of five runs (%)					
		1	2	3	4	5	Avg.
1	Train	59.82	59.82	59.82	59.82	59.82	59.82
	Test	52.62	52.62	52.62	52.62	52.62	52.62
5	Train	89.19	85.59	89.01	86.49	85.62	87.17
	Test	82.64	75.59	81.56	78.48	75.77	78.81
10	Train	89.91	85.95	92.79	86.31	91.53	89.30
	Test	80.65	75.05	84.81	77.58	81.37	79.89
20	Train	96.22	88.11	95.86	95.32	90.81	93.26
	Test	87.88	80.11	86.62	86.08	81.74	84.49
30	Train	85.77	84.14	95.68	88.65	86.13	88.07
	Test	79.20	77.40	81.92	79.57	78.66	79.35
40	Train	74.41	74.59	74.23	73.87	73.87	74.19
	Test	68.90	69.08	68.90	68.54	68.72	68.83

Table 6. Hit rates of PCA-d2b_M (MLM chips detected by 2 bands: MLM+MBM) at 10 percent FA rate. The MLP has either 1, 5, 10, 20, 30, or 40 inputs plus a bias.

Number of inputs	Data type	Hit rates at 10% FA of five runs (%)					
		1	2	3	4	5	Avg.
1	Train	31.35	31.35	31.35	31.35	31.35	31.35
	Test	28.21	28.21	28.21	28.21	28.21	28.21
5	Train	87.39	89.37	87.03	87.57	86.85	87.64
	Test	82.62	82.64	82.82	81.92	81.56	82.32
10	Train	97.93	90.45	89.55	89.55	92.79	90.05
	Test	96.44	87.16	87.52	90.42	89.15	88.14
20	Train	91.71	91.71	98.02	92.79	91.89	93.22
	Test	89.51	89.33	93.49	91.14	88.25	90.34
30	Train	92.79	91.17	88.65	94.05	91.71	91.67
	Test	84.99	86.26	84.63	90.24	88.07	86.84
40	Train	78.74	79.64	78.92	79.10	79.10	79.10
	Test	76.13	76.31	76.13	75.95	76.31	76.17

Table 7. Hit rates of EST-d2b_M (MLM chips detected by 2 bands: MLM+MBL) at 10 percent FA rate. The MLP has either 1, 5, 10, 20, 30, or 40 inputs plus a bias.

Number of inputs	Data type	Hit rates at 10% FA of five runs (%)					
		1	2	3	4	5	Avg.
1	Train	59.10	59.10	59.10	59.10	59.10	59.10
	Test	54.97	54.97	54.97	54.97	54.97	54.97
5	Train	89.01	88.65	87.39	83.78	85.77	86.92
	Test	83.91	81.92	79.93	79.02	80.11	80.98
10	Train	90.81	93.33	87.21	88.29	90.27	89.98
	Test	87.34	86.80	81.19	83.00	86.98	85.06
20	Train	90.09	96.04	94.05	96.94	96.58	94.74
	Test	86.62	87.52	86.26	90.24	89.69	88.07
30	Train	91.53	92.43	89.91	91.53	97.84	92.65
	Test	81.74	82.10	81.19	79.39	87.52	82.39
40	Train	71.17	70.99	70.63	71.35	70.27	70.88
	Test	65.64	64.20	64.56	62.93	64.01	64.27

Table 8. Hit rates of PCA-mrg (merged 2 bands: LLMMBL+MLMLBM) at 10 percent FA rate. The MLP has either 1, 5, 10, 20, 30, or 40 inputs plus a bias.

Number Of inputs	Data type	Hit rates at 10% FA of five runs (%)					
		1	2	3	4	5	Avg.
1	Train	40.90	40.90	40.90	40.90	40.90	40.90
	Test	36.71	36.71	36.71	36.71	36.71	36.71
5	Train	92.43	91.89	91.71	93.33	91.71	92.21
	Test	87.16	84.99	86.08	90.24	86.80	87.05
10	Train	97.84	96.40	96.76	98.02	94.77	96.76
	Test	92.22	92.41	92.41	95.30	91.86	92.84
20	Train	98.20	97.84	99.10	96.94	99.10	98.24
	Test	94.39	92.95	93.85	93.31	92.95	93.49
30	Train	100.00	98.74	96.58	98.74	99.82	98.78
	Test	93.49	93.31	91.86	92.41	93.31	92.88
40	Train	94.95	99.82	94.95	96.76	97.48	96.79
	Test	87.16	93.67	89.87	91.86	91.14	90.74

Table 9. Hit rates of EST-mrg (merged 2 bands: LLMMBL+MLMLBM) at 10% false alarm rate. The MLP has either 1, 5, 10, 20, 30, or 40 inputs plus a bias.

Number of inputs	Data type	Hit rates at 10% FA of five runs (%)					
		1	2	3	4	5	Avg.
1	Train	62.35	62.35	62.35	62.35	62.35	62.35
	Test	56.24	56.24	56.24	56.24	56.24	56.24
5	Train	94.05	91.35	92.61	93.15	93.33	92.90
	Test	89.15	88.61	89.69	90.24	90.60	89.66
10	Train	98.20	97.30	97.84	96.04	95.68	97.01
	Test	91.50	92.59	93.31	92.59	92.77	92.55
20	Train	94.05	97.84	97.30	97.48	96.94	96.72
	Test	86.44	91.50	92.22	89.87	90.05	90.02
30	Train	96.04	95.32	95.50	94.41	94.95	95.24
	Test	86.80	88.97	86.98	85.35	88.79	87.38
40	Train	90.63	91.53	91.17	94.59	93.51	92.29
	Test	84.81	84.63	82.46	85.35	84.45	84.34

Table 10. Hit rates of PCA-d2b_LM (separate eigenvector sets, joint MLP) at 10 percent FA rate. The MLP has either 2, 10, 20, 30, or 40 inputs plus a bias.

Number of inputs	Data type	Hit rates at 10% FA of five runs (%)					
		1	2	3	4	5	Avg.
2	Train	44.14	42.70	44.86	44.86	44.50	44.21
	Test	37.79	37.43	39.06	38.16	39.06	38.30
10	Train	97.66	95.50	96.94	94.59	95.50	96.04
	Test	91.32	89.51	91.86	89.69	93.13	91.10
20	Train	95.14	96.76	97.48	96.94	98.02	96.87
	Test	92.22	92.59	94.21	94.03	93.49	93.31
30	Train	94.77	96.04	97.12	94.77	96.40	95.82
	Test	90.42	91.50	91.50	93.31	93.31	92.01
40	Train	83.60	83.78	84.32	83.60	83.24	83.71
	Test	79.93	81.01	80.47	79.93	80.11	80.29

Table 11. Hit rates of EST-d2b_LM (separate eigenvector sets, joint MLP) at 10 percent FA rate. The MLP has either 2, 10, 20, 30, or 40 inputs plus a bias.

Number of inputs	Data type	Hit rates at 10% FA of five runs (%)					
		1	2	3	4	5	Avg.
2	Train	62.52	62.52	62.88	62.16	62.88	62.59
	Test	57.50	57.50	57.32	57.32	57.50	57.43
10	Train	92.79	93.15	93.33	92.97	93.33	93.11
	Test	89.33	87.70	88.61	87.88	87.52	82.21
20	Train	94.41	96.76	95.86	96.58	95.68	95.85
	Test	87.52	92.22	89.33	89.87	86.62	89.12
30	Train	94.23	93.69	92.25	95.50	94.77	94.09
	Test	85.71	86.44	84.63	88.07	88.61	86.69
40	Train	94.23	96.94	97.48	97.66	94.95	96.25
	Test	84.99	88.07	89.33	88.61	88.61	87.92

e. Hardware Implementation of Image Fusion

A Reconfigurable Computing module has been developed [xi] which is capable of implementing the three-module, center-surround shunt processing (CSSP) color fusion algorithm in real time similar to the Waxman [xii] fusion algorithm. The goal of this process is to produce a single image enhanced in such a way as to present the relevant information content from the original images in a form that is easily and naturally interpreted by the viewer. Algorithms for combining two images range from simple linear approaches such as pixel averaging, to complicated approaches that combine the pixel data using nonlinear function of the two pixel values. Among the latter are techniques that use information in a local region around a given pixel to modulate parameters in the fusion function.

A class of fusion algorithms also attempts to generate a false color image from two grayscale images. The three-module CSSP fusion algorithm was chosen for this hardware implementation based upon subjective evaluation of the simulation results. This

algorithm seemed to perform well and the false color enhancement provided a useful conduit for enhanced information content. Figure 18 shows a block diagram of the three-processor fusion algorithm and figure 19 shows an example of two-color IR fusion using this algorithm. The color map has been tuned so that the lake appears as blue-green.

The development approach for the reconfigurable digital signal processor (RCDSP) was guided by twin needs: to develop a computing solution capable of performing 640 by 480 image fusion at 30 Hz frame rate and to develop an extensible, experimental platform suitable for exploration of numerous other applications. In order to meet these twin needs, we undertook a study of several candidate algorithms to determine computational complexity and suitability for implementation. The algorithm chosen was the center-surround shunt processing image fusion algorithm shown schematically in figure 18.

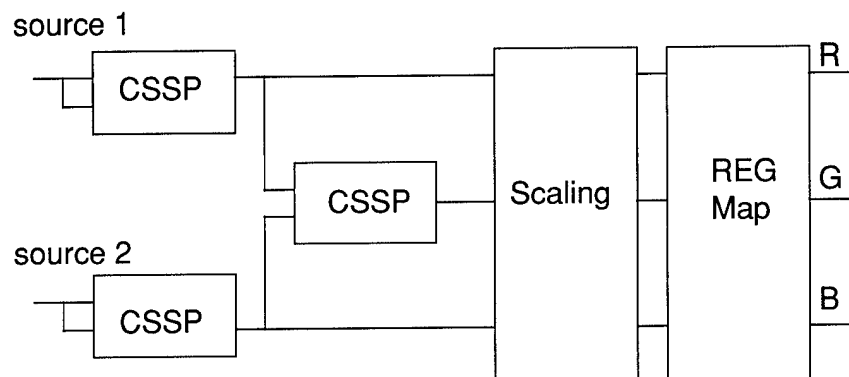


Figure 18. Block diagram of the three-processor, CSSP fusion algorithm.



Figure 19. The left image was taken with a cooled MWIR sensor, the center image was acquired with an uncooled LWIR sensor and the right is the result of processing with three CSSP to produce a false color enhanced image.

The center-surround shunt processing core operation is defined by eq (17). Design trade-offs associated with casting this equation into a form amenable to field-programmable gate array (FPGA) implementation will be described.

$$x_i = \left(\frac{B \cdot C \cdot G_{in} - D \cdot G_{out}}{A + (C \cdot G_{in} + G_{out}) * I} \right) * I \quad (17)$$

where:

x_i is the value of the i th pixel of the image.

I is the input image.

A, B, C and D are constants,

G_{in} is a Gaussian weighted mask for the central element, (usually set to unity) which gives the center element as the pixel.

G_{out} is a Gaussian weighted mask for the pixels surrounding the center pixel,

$*$ is the convolution operator.

The terms in and out refer to the center and the surround regions, respectively.

For the case where there are two different input images to the center-surround shunt processor, eq (17) becomes:

$$x_i = \frac{B \cdot C \cdot G_{in} * I_{in} - D \cdot G_{out} * I_{out}}{A + C \cdot G_{in} * I_{in} + G_{out} * I_{out}} \quad (18)$$

The convolutional kernels G_{in} and G_{out} are defined to be Gaussian and are therefore separable. This is taken advantage of in the FPGA implementation by performing row and then column Gaussian filtering with one-dimensional filters and performing the corner turn in an external RAM (random access memory) bank. The one-dimensional Gaussian filter is implemented as a cascade of first-order filters with coefficients of [1, 1]. Each of these small filters requires one add. The two-dimensional convolution takes $2N$ adds per output pixel. This is in contrast to the $O(N^2)$ multiplications (or additions) for the straightforward approach. The $2N$ additions cannot be parallelized so this implementation automatically introduces a one frame latency to the calculation but allows for more flexibility in determining the appropriate kernel size. The faster implementation creates row buffers inside the FPGA but this quickly becomes prohibitive for large kernels or large images.

By beginning the analysis with the more general case of eq (18), it is possible to determine the worst case computational complexity. As can be seen, the operations are two convolutions, 3 multiplies, one divide and 3 adds. The convolution can be expanded into $2N$ adds, 4 adds for bounds checking and limiting, and a scaling operation equivalent to 5 adds. The following are assumed:

- Image size is 640 by 480 pixels
- Frame rate is 30 frames/s
- Pixels require 16-Bit words
- Convolutional kernels are 9 by 9

For the choice of Xilinx 4000 series devices, one 16-bit adder requires 9 configurable logic blocks (CLBs) and a 16-bit multiply or divide requires 136 CLBs (based upon sizing estimation). The total CLB count for one center-surround shunt processor is $2 \cdot 9 \cdot (2 \cdot 9 + 9) + 3 \cdot 136 + 136 + 3 \cdot (9) = 1057$. The total for three center-surround shunt processors, not counting external interfaces is 3171 CLBs. The Xilinx 4085XL device has 3136 CLBs. Fortunately, in actual operation, the general case for the center-surround shunt processor is never implemented. In order to preserve image clarity and detail, the center convolutional kernel, G_{in} , is set to 1 by 1 so that no smoothing takes place. This essentially removes 243 CLBs from the total for one center-surround shunt processor. In addition, since the coefficients A, B, C and D are small numbers, the multipliers can be reduced to simple scaling (5 adder equivalent) resulting in $1 \cdot 9 \cdot (2 \cdot 9 + 9) + 3 \cdot 45 + 136 + 3 \cdot (9) = 541$ or 1623 CLBs, not counting external interfaces. This design can be implemented with reasonable confidence in the Xilinx 4085XL device chosen for the hardware.

For the case described above with an image size of 640 by 480 at 30 frames/s, the total number of 16-bit equivalent additions can be determined. The data rate is 9216000 pixels/s. For simplicity, take the incoming data rate to be 10 million pixels/s. If the divide operation is equivalent to 16 add operations, then each center-surround shunt processor consists of approximately 60 16-bit add equivalent operations per pixel. For three processors, the aggregate operation count is 1.8 Billion 16-bit add equivalent operations per second.

As previously mentioned, two corner-turn memories are required for each center-surround shunt processor. These are implemented as a virtual ping-pong buffer, one corner-turn to one RAM bank. This requirement of the algorithmic implementation placed a lower bound of six independent RAM banks on the hardware design. It also required that the RAM banks and the control circuitry operate at twice the incoming data rate, in this case 20 MHz, in order to support the virtual ping-pong structure. There are eight independent RAM banks on the RCDSP card, each of which is 1 Meg. by 16 bits, 15 ns access time. The minimum size required by the algorithm is 614400 16-bit locations. The total required memory bandwidth is 240 Mbytes/s. The total available memory bandwidth, assuming 40 MHz memory interface operation, is 640 Mbytes/s.

In addition to a 32-bit data path to the ADSP 21060 on the Alex Computer System PAC 509 card, the RCDSP supports 83 user I/O. Assuming 50 MHz. operation, the user input/output (I/O) alone provides over 500 Mbytes/s of I/O. The 32-bit link to the ADSP 21060 supports burst rates of up to 160 Mbytes/s.

A small, high performance FPGA-based computing module has been designed to implement a variety of signal processing algorithms. This FPGA card is coupled with a SHARC 21060-based processing card to create the RCDSP processing module. The three processor center-surround shunt two-color image fusion algorithm has been chosen as the first algorithm to be mapped to the RCDSP although several other algorithms were analyzed and their requirements considered in the design of the RCDSP. The RCDSP was demonstrated using archived image data in 1998. We expect to demonstrate this system with live dual-band imagery late in 1999 or early in 2000.

Current Status and Future Plans

As stated in the introduction, the ultimate goal of the MDSS effort is to obtain the imagery in the two infrared spectral bands from a single FPA. Early in 1999, BAE Systems (formerly Sanders, A Lockheed Martin Company) demonstrated a dual-band 256 by 256 focal plane array using QWIP technology [xiii]. Laboratory measurements show that the noise-equivalent temperature difference is 0.03 °C for both MWIR and LWIR bands at an operating temperature of 61 K. The detailed results of laboratory tests done on this FPA will be presented elsewhere [xiv].

Figure 20 shows an image obtained with the QWIP dual-band FPA. The left-hand image is LWIR and the right-hand image is MWIR. The man is holding a glass filter in front of a lit butane lighter. The filter is partially transparent in the MWIR and so the flame is visible in the MWIR image. The filter is completely opaque in the LWIR making the flame nearly invisible in the LWIR image. The entire plume is seen much better in the MWIR image than in the LWIR image. In addition, the reflection of the flame is seen on the man's hand in the MWIR image but not in the LWIR image. This behavior is expected because hot objects are known to be more visible in the MWIR and the MWIR is known to have a significant reflective component.

Figure 21 shows the results of the application of the image fusion algorithm discussed above on the images from figure 20. The flame and its reflection are seen as shades of cyan in the fused image because they were more prominent in the MWIR image. The man's skin appears red because it radiates more strongly in the LWIR.

LWIR

MWIR

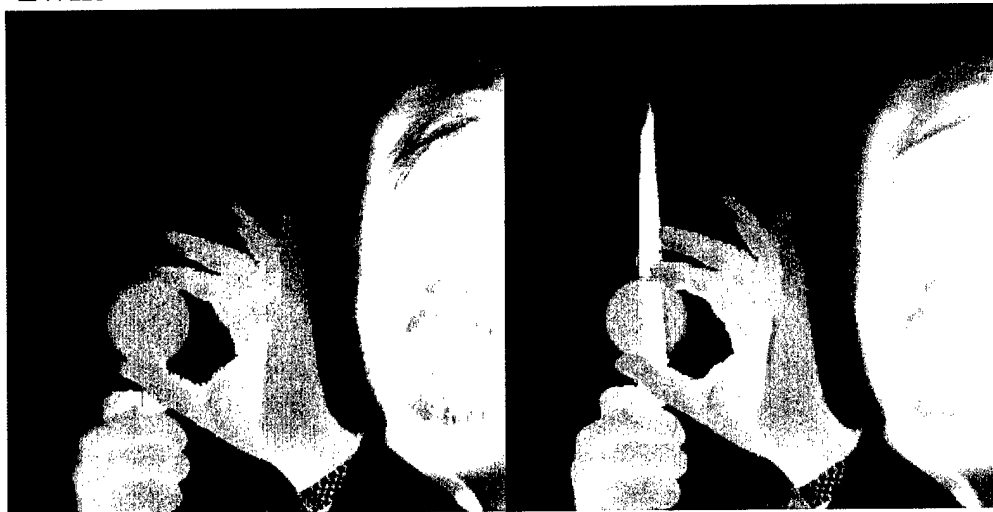


Figure 20. Image of a man holding a glass filter in front of a lit butane lighter taken with the QWIP dual-band FPA. In the left image (LWIR) the filter is opaque and the flame is not seen. In the right image (MWIR) the filter is partially transparent showing the flame. Both the flame and its reflection are much more prominent in the MWIR image.

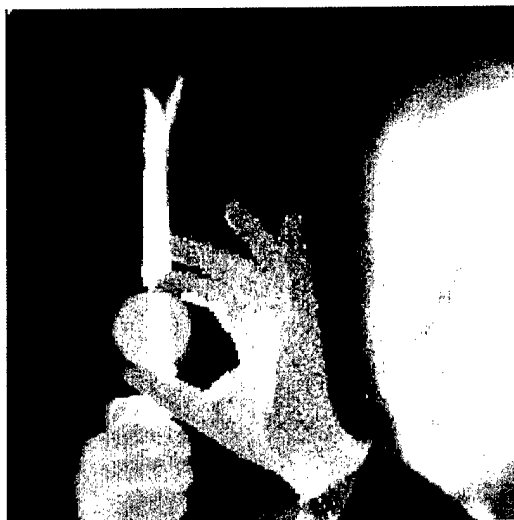


Figure 21. Result of image fusion on the images shown in fig. 20. The flame and its reflection emitted strongly in the MWIR and therefore are represented by shades of blue. The man's skin emits most strongly in the LWIR and is therefore mapped to shades of red.

It is our intention to take the dual-band FPA out into the field to gather data on targets under various ambient conditions including a wide range of obscurants. The dual-band FPA will be used in conjunction with image fusion algorithms. We hope that the data gathered in these tests will help to determine the best fusion algorithms and operating conditions for a conceptual MDSS system.

Summary and Conclusion

We have shown simultaneous IR imagery from the MWIR and LWIR bands taken at the MDSS field tests during summer, 1998. The imagery clearly shows the utility of dual-band IR imaging for (a) enhanced visibility through smoke (fig. 5), (b) greater operability in conditions of ground fog (fig. 6), and (c) enhanced visibility of objects not seen well in either band separately (fig. 9). In addition, we have shown that a color fusion algorithm can be used to map the information contained in the separate MWIR and LWIR images into a single image that can give the observer increased situational awareness. We have shown a path for implementing the image fusion in hardware at real-time frame rates. Finally, we have shown that the use of dual-band imagery can significantly reduce missed detections in and ATR for a fixed false alarm rate as compared with either LWIR or MWIR imagery alone.

Acknowledgments

We would like to acknowledge Major R. C. Deluca, S. Kennerly, and G. Green of ARL for invaluable help in setting up and carrying out the field tests. We gratefully acknowledge the management support of J. Pellegrino, D. Wilmot, H. Pollehn, and G. Sztankay at ARL and J. Ahearn at BAE Systems.

References

- [1] H. Pollehn and J. Ahearn, "Multi-Domain Smart Sensors," Proceedings of the SPIE, Infrared Technology and Applications XXV, Vol. 3698, Orlando, FL (1999).
- [2] D. Scribner, J. Schuler, P. Warren, M. Satyshur, M. Kruer, "Infrared Color Vision: Separating Objects from Backgrounds", Proceedings of the SPIE - Infrared Detectors and Focal Plane Arrays V, Vol. 3379, (1998).
- [3] C. Dwan and S. Der, "A Neural Net Based Target Detection System for FLIR Imagery," Proceedings of the Second Federated Laboratory Symposium on Advanced Sensors, pp. 183-187, (1998).
- [4] R.C. Gonzalez and R.E. Woods, Digital Image Processing, Addison-Wesley Publishing, New York, (1992).
- [5] S. Fahlman, "Faster Learning Variations on Back-Propagation: An Empirical Study," Proceedings of the 1988 Connectionist Models Summer School, Morgan Kaufmann, pp. 38-51 (1998).
- [6] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, Numerical Recipes in C, Second Edition, Cambridge University Press, New York (1992).
- [7] D. Torrieri, "A Linear Transformation that Simplifies and Improves Neural Network Classifiers," Proceedings of the International Conference on Neural Networks, 3, pp. 1738-1743 (1996).
- [8] G.L. Plett, T. Doi, and D. Torrieri, "Mine Detection Using Scattering Parameters and an Artificial Neural Network," IEEE Transactions on Neural Networks 8, no. 6, pp. 1456-1467 (1997).
- [9] R. Anand, K. Mehrotra, C. Mohan, and S. Ranka, "An Improved Algorithm for Neural Network Classification of Imbalanced Training Sets," IEEE Transactions on Neural Networks 4, no. 6, pp. 962-969 (1996).
- [10] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan College Publishing, New York, (1994).
- [11] A. Castillo, D. Compagna M. Falco, and A. Filipov, "Re-configurable Digital Signal Processor with Application Using a Waxman-like Fusion Algorithm," Proceedings of the Third Federated Laboratory Symposium on Advanced Sensors, pp. 163-167 (1999).
- [12] A.M. Waxman, D.A. Fay, A.N. Gove, M.C. Siebert, and J.P. Racamato, "Method and Apparatus for Generating a Synthetic Image by the Fusion of Signals Representative of Different Views of the Same Scene," U. S. Patent Application 08/332,696, submitted 11/1/94, legal claims approved 11/95.

- [13] P. Uppal, M. Sundaram, A. Reisinger, S. Wang, M. Taylor, T. Faska, J. Little, W. Beck, A. Goldberg, and S. Kennerly, "Status of Two-color LWIR/MWIR QWIP Focal Plane Arrays," Proceedings of the Third Federated Laboratory Symposium on Advanced Sensors, pp. 29-32, 1999.
- [14] A. Goldberg, S. Wang, M. Sundaram, P. Uppal, M. Winn, G. Milne and M. Stevens, "Dual Band MWIR/LWIR Focal Plane Array Test Results," Proc. 1999 IRIS Specialty Group Meeting on Detectors (in print).

List of Acronyms

APC	Armored personnel carrier
ARL	Army Research Laboratory
ATR	Automatic target recognition
CLB	Configurable Logic Block
CR	Clutter Rejection
CSSP	Center surround shunt processor
EST	Eigenspace Separation Transformation
FA	False alarm
FOV	Field of view
FPA	Focal plane array
FLIR	Forward looking infrared
FPGA	Field programmable gate array
HC	Hexachloroethane
IR	Infrared
LWIR	Long wavelength infrared
MDSS	Multi-domain smart sensor
MLP	Multi-layer perception
MWIR	Medium wavelength infrared
PC	Principal component
PCA	Principal component analysis
RCDSP	Reconfigurable Digital Signal Processor

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE June 2002		3. REPORT TYPE AND DATES COVERED Final, 6/1/98-5/1/99
4. TITLE AND SUBTITLE Analysis of Dual-Band Infrared Imagery from the Multidomain Smart Sensor Field Test			5. FUNDING NUMBERS DA PR: AH94 PE: 62705A	
6. AUTHOR(S) A. Goldberg, T. Fisher, S. Kennerly, S. Der, A. Chan, M. Lander (ARL), C. Garvin, S. Wang, M. Falco, D. Campagna, and A. Costillo (Sanders)				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory Attn: AMSRL- SE-EE email: arniec@arl.army.mil 2800 Powder Mill Road Adelphi, MD 20783-1197			8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-996	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Laboratory 2800 Powder Mill Road Adelphi, MD 20783-1197			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES ARL PR: 9NE6CC AMS code: 622705.H9411				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) Included is dual-band infrared image data collected as part of the Multi-domain Smart Sensor effort at the U. S. Army Research Laboratory. The ultimate goal of this effort is to produce large format, staring focal plane arrays that are able to see the battlefield in both the 3 to 5 m (midwave infrared) and 8 to 12 m (longwave infrared) atmospheric transmission windows. The image data were collected using separate boresighted cameras with equal pixel formats and fields of view during field tests that were conducted during the summer of 1998. This work shows a number of scenarios under which the imagery from one band is superior to that from the other band and various image fusion techniques that help to enhance the visibility of targets. Discussed is a technique for using computer hardware to do the image fusion in real time as well as results of the application of aided target recognition algorithms to the data.				
14. SUBJECT TERMS Infrared imagery dual-band analysis			15. NUMBER OF PAGES	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR	